Simfit

MANOVA examples

From the main SɪᴍFɪT menu choose [Statistcs], [Multivariate], then [MANOVA], noting that several test files named `manova1.tfk` are provided for $k = 1$ to $k = 5$. It is important to realize that the first column in all data sets provided for MANOVA analysis must have the group numbers as successive integers in column 1 in nondecreasing order, with further columns for observations.

MANOVA example 1. Testing for equality of all means

Example 1 describes the results from analyzing these data for three groups and two variables contained in test file `manova1.tf3`.

1	3	10
1	5	16
1	5	16
1	4	14
1	1	9
2	8	12
2	4	8
2	4	6
2	2	6
2	9	14
3	10	16
3	4	10
3	10	18
3	4	14
3	10	16

Column 1 is the group number (in nondecreasing order), while columns 2 and 3 are the observations.

If all groups have the same multivariate normal distribution, then estimates for the mean μ and covariance matrix Σ can be obtained from the overall sample statistics $\hat{\mu} = \bar{y}$ and $\hat{\Sigma}$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n_i} y_{ij}$$

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^{g} \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu})(y_{ij} - \hat{\mu})^T$$

obtained by ignoring group means $\bar{y}_i$ and summing across all groups. Alternatively, the pooled between-groups B, within-groups W, and total sum of squares and products matrices T can be obtained along with the

within-groups covariance matrix S using the group mean estimates $\bar{y}_i$ as

$$B = \sum_{i=1}^{g} n_i(\bar{y}_i - \bar{y})(\bar{y}_i - \bar{y})^T$$

$$W = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)(y_{ij} - \bar{y}_i)^T$$

$$= \sum_{i=1}^{g} (n_i - 1)S_i$$

$$= (n - g)S$$

$$T = B + W$$

$$= (n - 1)\hat{\Sigma}.$$

This table is typical, and clearly strong differences between groups will be indicated if B is much larger than W.

Source of variation	d.f.	ssp matrix
Between groups	$g - 1$	B
Within groups	$n - g$	W
Total	$n - 1$	T

The usual likelihood ratio test statistic is Wilk's lambda defined as

$$\Lambda = \frac{|W|}{|B| + |W|}$$

but other statistics can also be defined as functions of the eigenvalues of BW^{-1}. Unlike B and W separately, the matrix BW^{-1} is not symmetric and positive definite but, if the m eigenvalues of BW^{-1} are θ_i, then Wilk's lambda, Roy's largest root R, the Lawley-Hotelling trace T, and the Pillai trace P can be defined as

$$\Lambda = \prod_{i=1}^{m} \frac{1}{1 + \theta_i}$$

$$R = \max(\theta_i)$$

$$T = \sum_{i=1}^{m} \theta_i$$

$$P = \sum_{i=1}^{m} \frac{\theta_i}{1 + \theta_i}.$$

The next table of results was obtained when `manova1.tf3` was analyzed, and the methods used to calculate the significance levels will then be outlined.

MANOVA H_0: all mean vectors are equal

Number of groups	3
Number of variables	2
Number of observations	15

Statistic	Value	Transform	$NDOF$	p	
Wilks lambda	0.1917	7.062	4, 22	0.0008	*Reject H_0 at 1%*
Roys largest root	2.801				
Lawley-Hotelling T	3.173	8.727	4, 11	0.0017	*Reject H_0 at 1%*
Pillais trace	1.008				

The next table indicates conditions on the number of groups g, variables m, and total number of observations n that lead to exact F variables for appropriate transforms of Wilk's Λ.

Parameters	F statistic	Degrees of freedom
$g = 2$, any m	$\dfrac{(2g - m - 1)(1 - \Lambda)}{m\Lambda}$	$m, 2g - m - 1$
$g = 3$, any m	$\dfrac{(3g - m - 2)(1 - \sqrt{\Lambda})}{m\sqrt{\Lambda}}$	$2m, 2(n - m - 2)$
$m = 1$, any g	$\dfrac{(n - g)(1 - \Lambda)}{(g - 1)\Lambda}$	$g - 1, n - g$
$m = 2$, any g	$\dfrac{(n - g - 1)(1 - \sqrt{\Lambda})}{(g - 1)\sqrt{\Lambda}}$	$2(g - 1), 2(n - g - 1)$

For other conditions the asymptotic expression

$$-\left(\frac{2n - 2 - m - g}{2}\right)\log\Lambda \sim F_{m,g-1}$$

is generally used. The Lawley-Hotelling trace is a generalized Hotelling's T_0^2 statistic, and so the null distribution of this can be approximated as follows.

Defining the degrees of freedom and multiplying factors α and β by

$$\nu_1 = g - 1$$
$$\nu_2 = n - g$$
$$\nu = \frac{m\nu_1(\nu_2 - m)}{\nu_1 + \nu_2 - m\nu_1 - 1}$$
$$\alpha = \frac{(\nu_2 - 1)(\nu_1 + \nu_2 - m - 1)}{(\nu_2 - m)(\nu_2 - m - 1)(\nu_2 - m - 3)}$$
$$\beta = \frac{m\nu_1}{\nu_2 - m + 1},$$

then the case $\nu > 0$ leads to the approximation

$$T \sim \beta F_{\nu, \nu_2 - m + 1},$$

otherwise the alternative approximation

$$T \sim \alpha\chi_f^2$$

is employed, where $f = m\nu_1/\{\alpha(\nu_2 - m - 1)\}$. The null distributions for Roy's largest root and Pillai's trace are more complicated to approximate, which is one reason why Wilk's Λ is the most widely used test statistic.

MANOVA example 2. Testing for equality of selected means

The next table resulted when groups 2 and 3 were tested for equality of selected means, another example of a Hotelling's T^2 test.

MANOVA H_0: selected group means are equal

First group	2 (5 cases)	
Second group	3 (5 cases)	
Number of observations	15 (to estimate CV)	
Number of variables	2	
Hotelling T^2	12.00	
Test statistic S	5.498	
Numerator DOF	2	
Denominator DOF	11	
$P(F \geq S)$	0.0221	*Reject H_0 at 5% significance level*

MANOVA H_0: selected group means are equal

First group	2 (5 cases)	
Second group	3 (5 cases)	
Number of observations	10 (to estimate CV)	
Number of variables	2	
Hotelling T^2	15.18	
Test statistic S	6.640	
Numerator DOF	2	
Denominator DOF	7	
$P(F \geq S)$	0.0242	*Reject H_0 at 5% significance level*

The first result uses the difference vector $d_{2,3}$ between the means estimated from groups 2 and 3 with the matrix $W = (n - g)S$ estimated using the pooled sum of squares and products matrix to calculate and test T^2 according to

$$T^2 = \left(\frac{(n - g)n_2 n_3}{n_2 + n_3} \right) d_{2,3}^T W^{-1} d_{2,3}$$

$$\frac{n - g - m + 1}{m(n - g)} T^2 \sim F_{m, n-g-m+1},$$

while the second result uses the data from samples 2 and 3 as if they were the only groups as follows

$$S_{2,3} = \frac{(n_2 - 1)S_2 + (n_3 - 1)S_3}{n_2 + n_3 - 2}$$

$$T^2 = \left(\frac{n_2 n_3}{n_2 + n_3} \right) d_{2,3}^T S_{2,3}^{-1} d_{2,3}$$

$$\frac{n_2 + n_3 - m - 1}{m(n_2 + n_3 - 2)} T^2 \sim F_{m, n_2+n_3-m-1}.$$

The first method could be used if all covariance matrices are equal (see next) but the second might be preferred if it was only likely that the selected covariance matrices were identical.

MANOVA example 3. Testing for equality of all covariance matrices

---

The next data set in `manova1.tf2` has three groups for three types of Cushing's syndrome, the variables are logarithms of urinary excretion rates (mg/hr) for two steroid metabolites, and the table below the data shows the results from testing that the within-group variance-covariance matrices are equal.

```
1    1.1314    2.4596
1    1.0986    0.2624
1    0.6419   -2.3026
1    1.3350   -3.2189
1    1.4110    0.0953
1    0.6419   -0.9163
2    2.1163    0.0000
2    1.3350   -1.6094
2    1.3610   -0.5108
2    2.0541    0.1823
2    2.2083   -0.5108
2    2.7344    1.2809
2    2.0412    0.4700
2    1.8718   -0.9163
2    1.7405   -0.9163
2    2.6101    0.4700
3    2.3224    1.8563
3    2.2192    2.0669
3    2.2618    1.1314
3    3.9853    0.9163
3    2.7600    2.0281
```

MANOVA H_0: all covariance matrices are equal

Number of groups	3	
Number of observations	21	
Number of variables	2	
Test statistic C	19.24	
Degrees of freedom	6	
$P(\chi^2 \geq C)$	0.0038	*Reject H_0 at 1% significance level*

These results refer to using Box's test to analyze `manova1.tf2` for equality of covariance matrices. This depends on the likelihood ratio test statistic C defined by

$$C = M \left\{ (n - g) \log |S| - \sum_{i=1}^{g} (n_i - 1) \log |S_i| \right\},$$

where the multiplying factor M is

$$M = 1 - \frac{2m^2 + 3m - 1}{6(m + 1)(g - 1)} \left(\sum_{i=1}^{g} \frac{1}{n_i - 1} - \frac{1}{n - g} \right)$$

and, for large n, C is approximately distributed as χ^2 with $m(m + 1)(g - 1)/2$ degrees of freedom. Just as tests for equality of variances are not very robust, this test should be used with caution, and then only with large samples, i.e. $n_i \gg m$.

MANOVA example 4. Profile analysis

Test file `manova1.tf1` has two groups and five variables as follows,

```
1  11  18  15  18  15
1  33  27  31  21  17
1  20  28  27  23  19
1  18  26  18  18   9
1  22  23  22  16  10
2  18  17  20  18  18
2  31  24  31  26  20
2  14  16  17  20  17
2  25  24  31  26  18
2  36  28  24  26  29
```

while the following table shows the results of statistical analysis using the profile option.

MANOVA H_0: selected group profiles are equal		
First group	1 (5 cases)	
Second group	2 (5 cases)	
Number of observations	10 (to estimate CV)	
Number of variables	5	
Hotelling T^2	35.65	
Test statistic S	5.570	
Numerator DOF	4	
Denominator DOF	5	
$P(F \geq S)$	0.0438	*Reject H_0 at 5% significance level*

The next figure illustrates the results from plotting the group means from `manova1.tf1` using the profile analysis option, noting that error bars are not added as a multivariate distribution is assumed,

MANOVA Profile Analysis
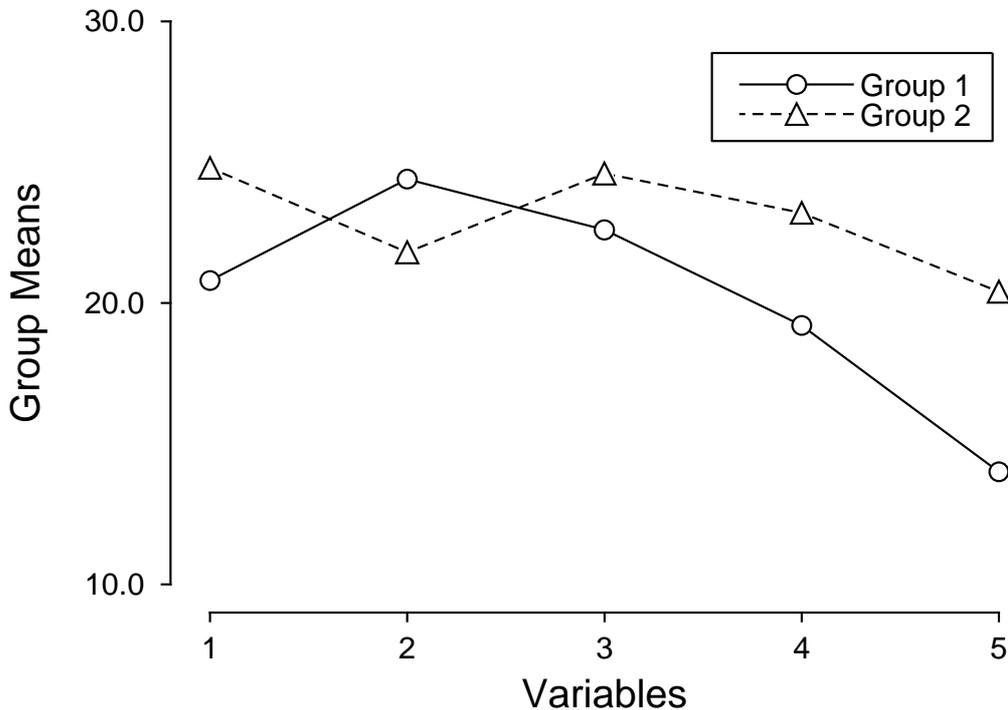
Profile analysis attempts to explore a common question that often arises in repeated measurements ANOVA namely, can two profiles be regarded as parallel. This amounts to testing if the sequential differences between

adjacent means for groups i and j are equal, that is, if the slopes between adjacent treatments are constant across the two groups, so that the two profiles represent a common shape.

To do this, we first define the $m-1$ by m transformation matrix K by

$$K = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & 0 & \dots \\ 0 & 0 & 1 & -1 & \dots & \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Then a Hotelling's T^2 test is conducted using the pooled estimate for the covariance matrix $S_{ij} = [(n_i - 1)S_i + (n_j - 1)S_j]/(n_i + n_j - 2)$ and mean difference vector $d_{ij} = \bar{y}_i - \bar{y}_j$ according to

$$T^2 = \left(\frac{n_i n_j}{n_i + n_j} \right) (Kd_{ij})^T (KS_{ij}K^T)^{-1}(Kd_{ij})$$

and comparing the transformed statistic

$$\frac{n_i + n_j - m}{(n_i + n_j - 2)(m-1)} T^2 \sim F_{m-1, n_1 + n_2 - m}$$

to the corresponding F distribution.

Clearly, from the above table, the profiles are not parallel for the data in test file `manova1.tf1`.